Tools for Exploring Functional Data

Pang Du

Department of Statistics
Virginia Tech

Overview

- Summary statistics for functional data
- Anatomy of a function
- Phase-plane plots of periodic effects

Notation

- Vector: $\mathbf{x}$. Matrix: $\mathbf{X}$. Transpose: $\mathbf{X}'$.
  Scalar function: $x(\cdot)$. Vector function: $\mathbf{x}(\cdot)$.
- m-th derivative: $D^m x$. Indefinite integral: $D^{-1}x$.
  And $D^0 x = x$.
- Inner product of functions: $\langle x, y \rangle = \int_{\mathcal{T}} x(t)y(t)dt$.
  Norm: $\|x\|^2 = \langle x, x \rangle = \int_{\mathcal{T}} x^2(t)dt$.
  Here $\mathcal{T}$ is the domain of t.
- Functional composition: $(x \circ h)(t) = x(h(t))$.

Univariate data

- Functional mean: $\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t)$.
- Functional variance:

$$\mathrm{var}_X(t) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i(t) - \bar{x}(t))^2.$$

- Functional covariance:

$$\mathrm{cov}_X(t_1, t_2) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i(t_1) - \bar{x}(t_1))(x_i(t_2) - \bar{x}(t_2)).$$

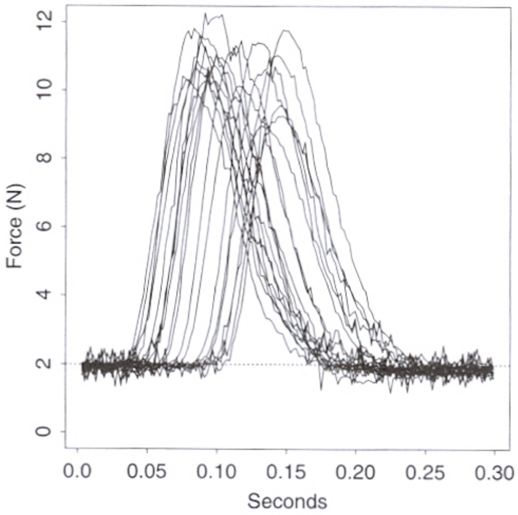
- Functional correlation:

$$\mathrm{corr}_X(t_1, t_2) = \frac{\mathrm{cov}_X(t_1, t_2)}{\sqrt{\mathrm{var}_X(t_1)\mathrm{var}_X(t_2)}}.$$

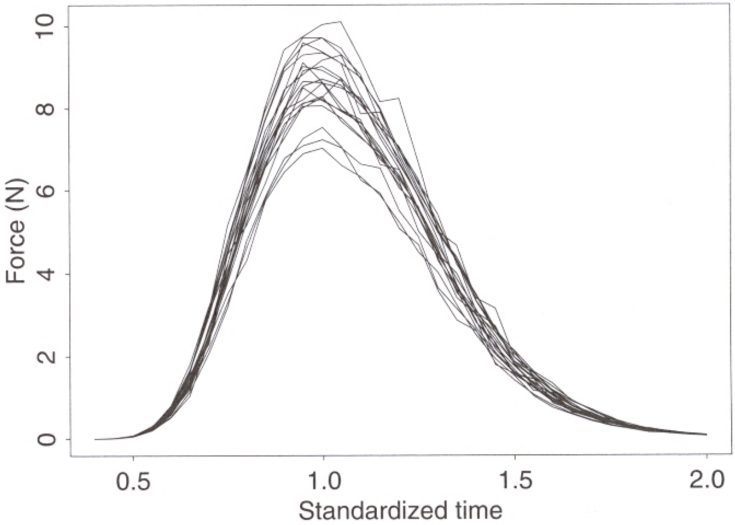
The pinch force data

- Reference: Ramsay, Wang and Flanagan (1995). "A functional data analysis of the pinch force of human fingers". *Applied Statistics* **44**: 17-30.
- The data consisted of $N = 20$ records of the force exerted on a meter during a brief pinch by the thumb and forefinger.
- The subject was required to maintain a certain background force on a force meter and then to squeeze the meter aiming at a specified maximum value, returning afterwards to the background level.
- The purpose of the biomechanical experiment was to study the neurophysiology of the thumb-forefinger muscle group.

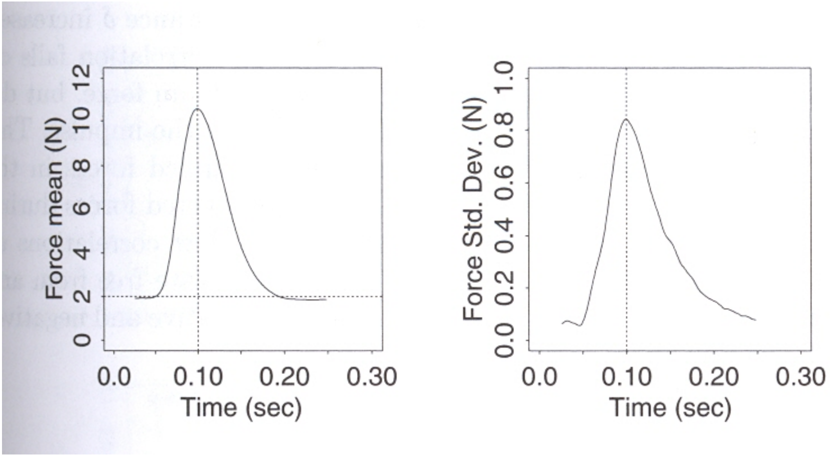
Pinch force data



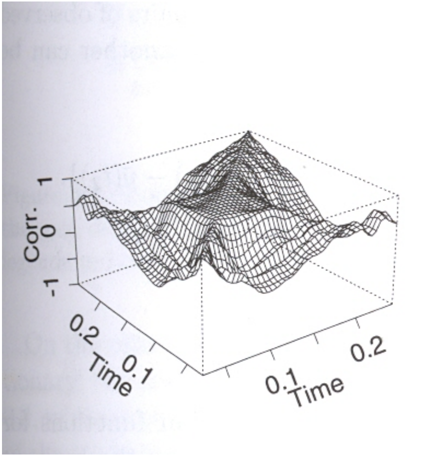
Pinch force data after registration

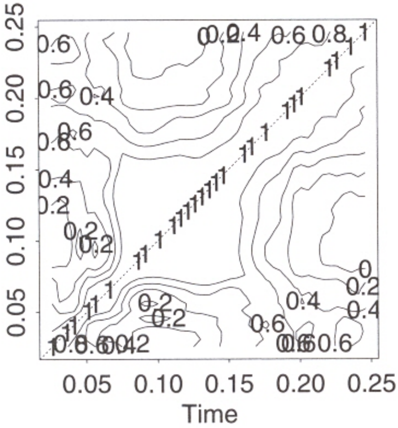


Mean & Std. Dev. of pinch force data



Correlation plot of pinch force data





Findings from plots of pinch force data

- Mean force plot has a shape resembling a log-normal density.
- Standard deviation plot shows the most variation at the moment of the pinch.
- Correlations decrease slowly around the moment of the pinch ($t = 0.1$) and more rapidly at both ends of the recording.

Bivariate data

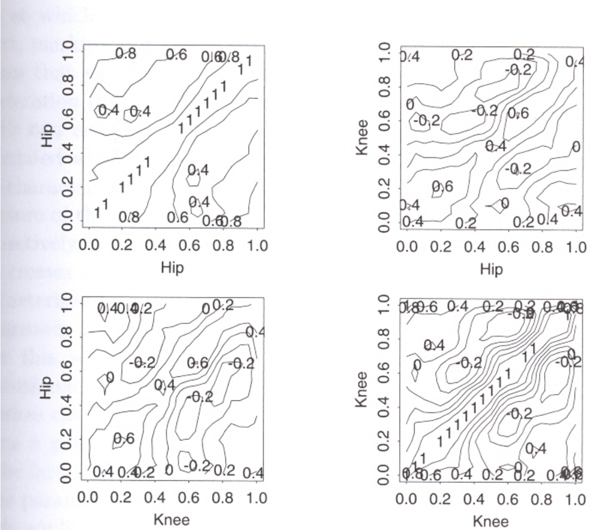
- Cross-covariance:

$$\text{cov}_{X,Y}(t_1, t_2) = \frac{1}{N-1} \sum_{i=1}^{N} (x_i(t_1) - \bar{x}(t_1))(y_i(t_2) - \bar{y}(t_2)).$$

- Cross correlation:

$$\text{corr}_X(t_1, t_2) = \frac{\text{cov}_{X,Y}(t_1, t_2)}{\sqrt{\text{var}_X(t_1)\text{var}_Y(t_2)}}.$$

Correlation & cross-correlation of gait data



Findings from plots of gait data

- Hip-Hip correlation:
  - Positive correlation throughout the range.
  - contours parallel to diagonal (approx. a function of $|t_1 - t_2|$ and stationary variation).
- Knee-Knee correlation:
  - Positive correlation at both ends of the cycle.
  - negative correlation in the middle of the cycle.
- Knees are more affected by external factors such as heel strike, whereas the hip acts under much more even muscular control throughout the cycle.
- Hip-Knee cross-correlation:
  - positive and most strongly correlated on the main diagonal.
  - different behaviors in cross-correlation at different points of the cycle.

Functional features and dimensionality

- Functional features:
  - Peaks and valleys: growth spurts.
  - Crossing of specific levels: age of zero acceleration (peak growth velocity).
  - Levels: function values considered to be significant.
- Dimensionality of functional features:
  - Peaks and valleys: 3D (location, amplitude, width), like a parabola.
  - Crossings: 2D (location, level), like a line.
  - Levels: 1D (value of level), like a point.
- Functional dimensionality:
  - sum of the numbers of pieces of information that are required to define all the features of a function.
  - white noise and Brownian motion functions are of “infinite dimensions”.
  - smooth functions are of lower dimensions.

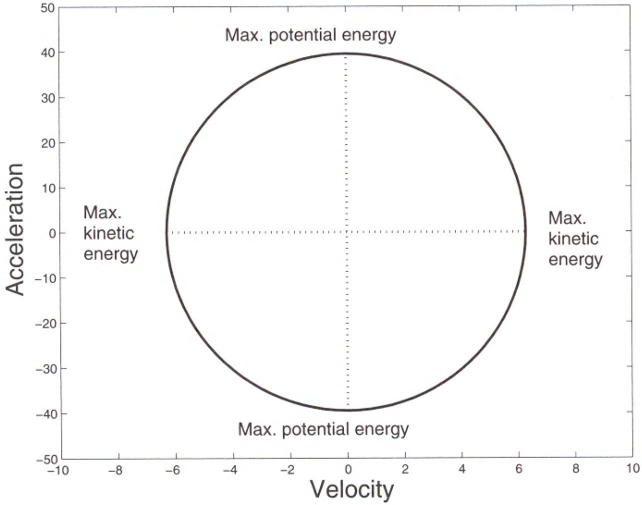
"Energy" of a functional variable

- Many functional variables are governed by "*energy*":
  - Stock market: effort required to move around money and information.
  - Child's growth: consumed calories.
  - Rainfall: water accumulation in clouds.
- A function is described by its dimensionality and amplitude, both of which require energy to produce.

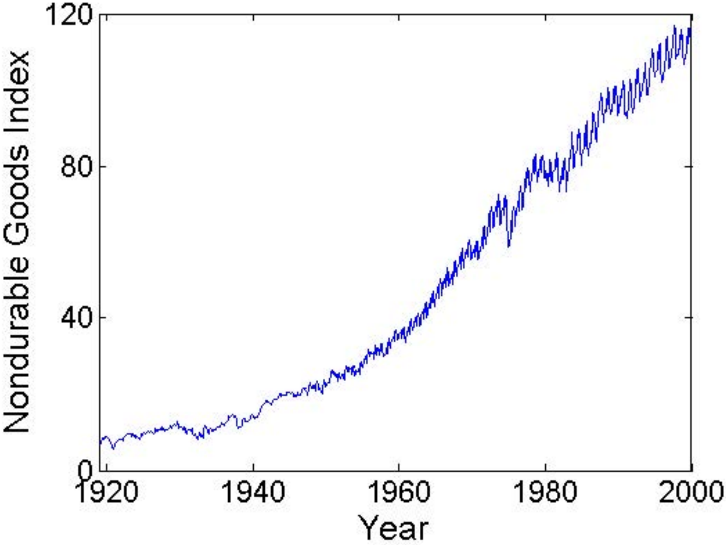
A simple physic system

- A suspended string bouncing in the vacuum with a period of one time unit.
- Start recording the vertical position $x(t)$ of the end of the string when it arrives at the equilibrium position (i.e., $x(0) = 0$).
- $x(t) = \sin(2\pi t)$ is a *harmonic function*.
- Oscillation of the string reflects the exchange of energy between two states: *potential* and *kinetic*.

Phase-plane plot of harmonic function



U.S. nondurable goods index



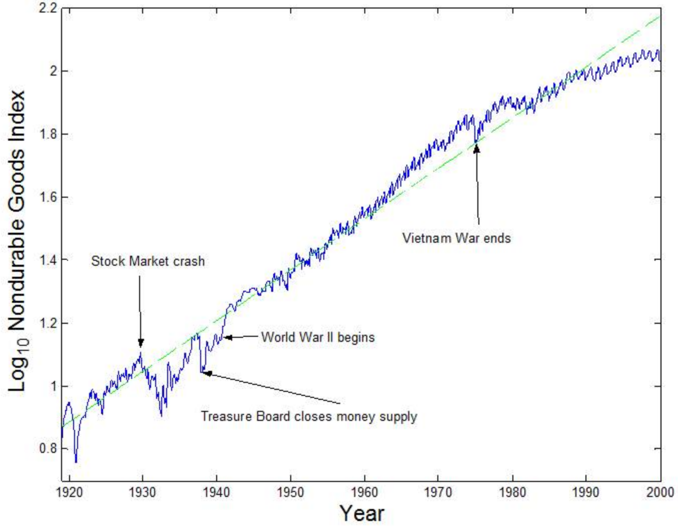
Background

- Nondurable goods last less than two years: Food, clothing, cigarettes, alcohol, but not personal computers!!
- The nondurable goods manufacturing index is an indicator of the economics of everyday life.
- The index has been published monthly by the US Federal Reserve Board since 1919.
- It complements the durable goods manufacturing index.

Our interests

- Look at important events.
- Examine the overall trend in the index.
- Have a look at the annual or seasonal behavior of the index.
- Understand how the seasonal behavior changes over the years and with specific events.

Log nondurable goods index



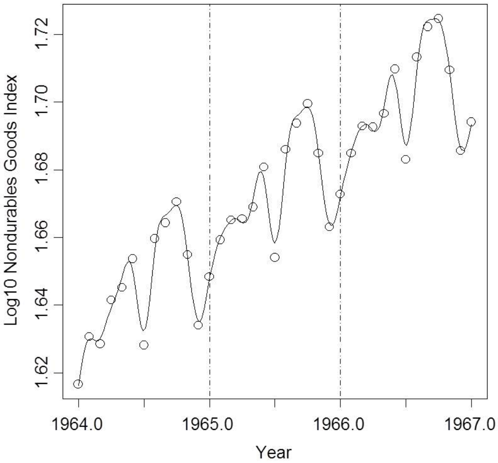
Events and trends

- Short term:
  - 1929: stock market crash.
  - 1937 restriction of money supply.
  - 1975 end of OPEC oil crisis and Vietnam War.
- Medium term:
  - 1929-1939: Great Depression.
  - 1939-1945: World War II.
  - Unusually rapid growth in 1960-1974.
  - Ususually slow growth in 1990-2000.
- Long term increase of 1.6% per year.

Evolution of seasonal trend

- We first focus on the years 1948 to 1999.
- We estimate long- and medium-term trend by spline smoothing.
- These estimates have knots too far apart to capture seasonal trend.
- We subtract this smooth trend to leave only seasonal trends.

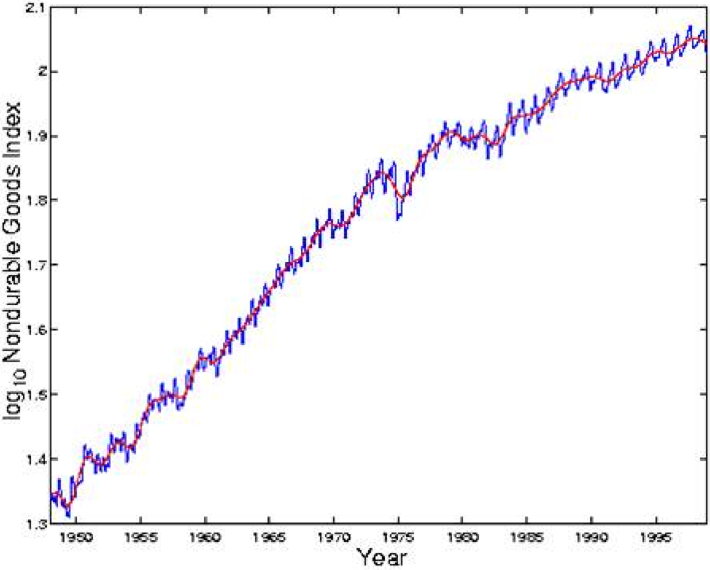
Three years of typical trend: 1964-66



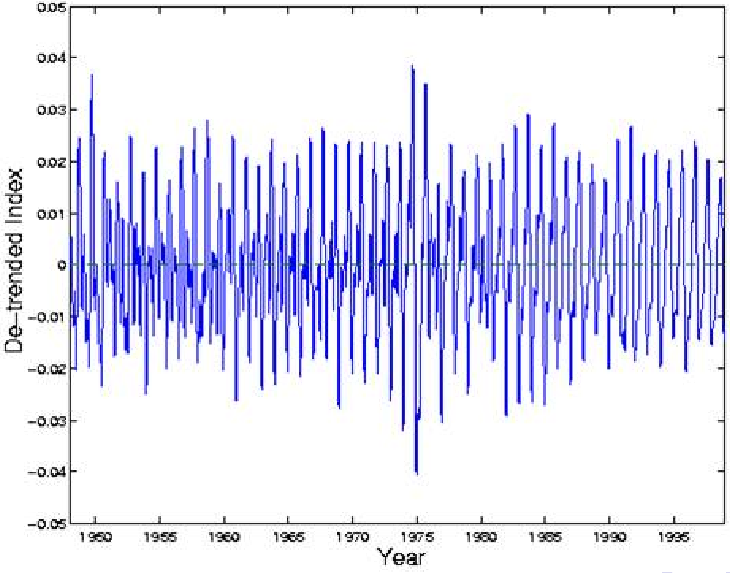
Evolution of seasonal trend

- Typically three peaks per year.
- The largest is in the fall, peaking at the beginning of October.
- The low point is mid-December.

Non-seasonal trend



Seasonal trend

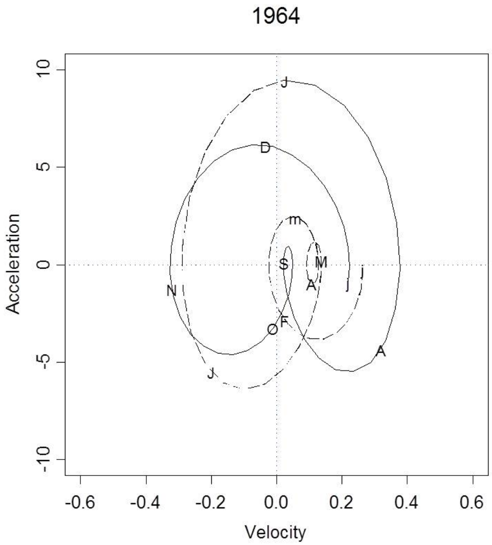


Phase-plane plots in manufacturing

"Energy transfer" in manufacturing process:

- Potential energy: available capital, human resources, raw material, and other resources that are at hand for the manufacture of nondurable goods.
- Kinetic energy: manufacturing process in full swing when these resources are moving along the assembly line and the goods are being shipped out of the factory door.
- Strong potential status: rate of change in production close to zero.
    - resources depleted.
    - production target achieved.
- Low potential and kinetic energy at periods of crisis: phase-plane curve close to zero.

Phase-plane plot of year 1964

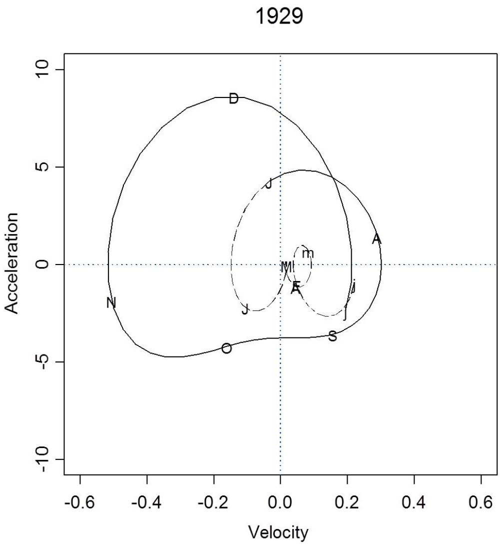


Seasonal trends in 1964

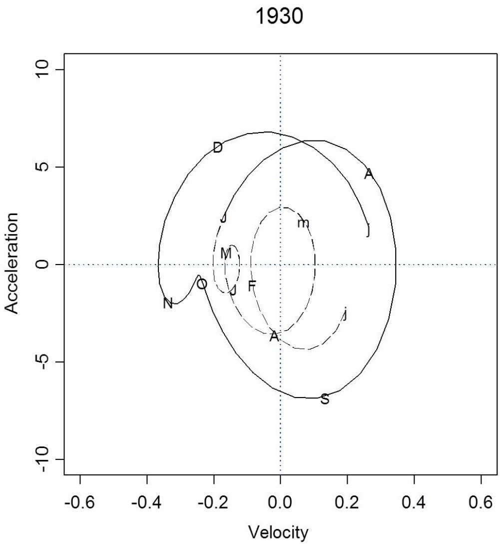
There are three large loops separated by two small loops or cusps:

- **Spring cycle**: mid-January into April.
- **Summer cycle**: May through August.
- **Fall cycle**: October through December.
- **Two cusps**: April to May & in September.

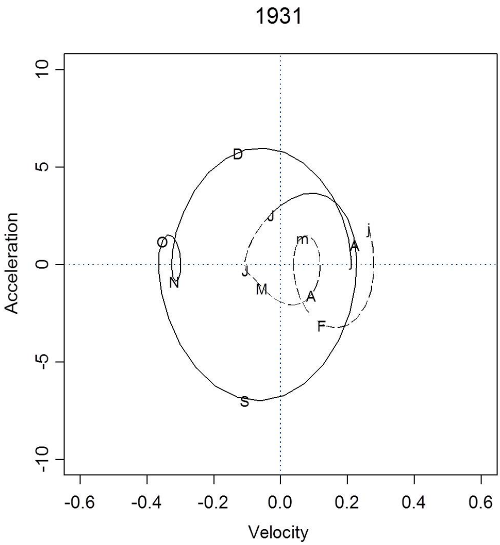
Phase-plane plot of year 1929



Phase-plane plot of year 1930



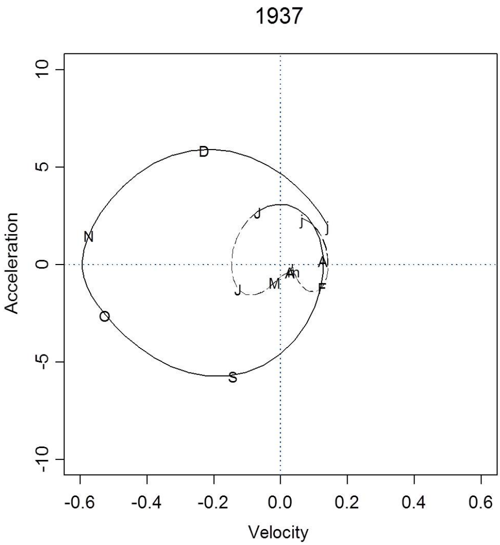
Phase-plane plot of year 1931



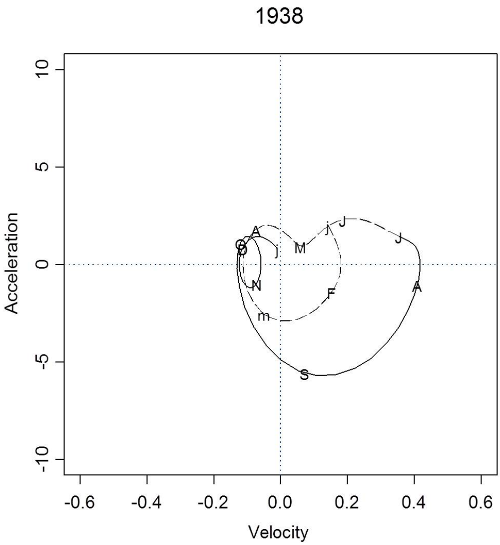
Years 1929-1931

- Year 1929: the stock market crash in October shows up as a large negative surge in velocity.
- Subsequent years nearly lose the fall production cycle, as people tighten their belts and spend less at Christmas.

Phase-plane plot of year 1937



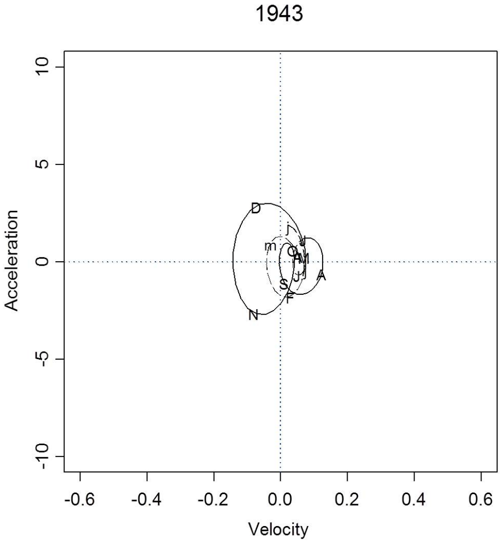
Phase-plane plot of year 1938



Years 1937-1938

- Small cycles at the beginning of 1937 were due to the Great Depression.
- The Treasury Board, fearing that the economy was becoming overheated again, clamped down on the money supply in the fall. The effect was catastrophic, and nearly wiped out the fall cycle.
- This new crash was even more dramatic than that of 1929, also wiping out the spring and summer cycles in 1938.
- But this was forgotten because of the outbreak of World War II...

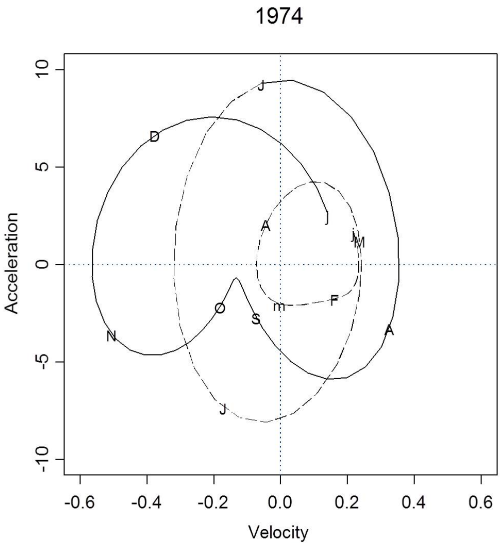
Phase-plane plot of year 1943



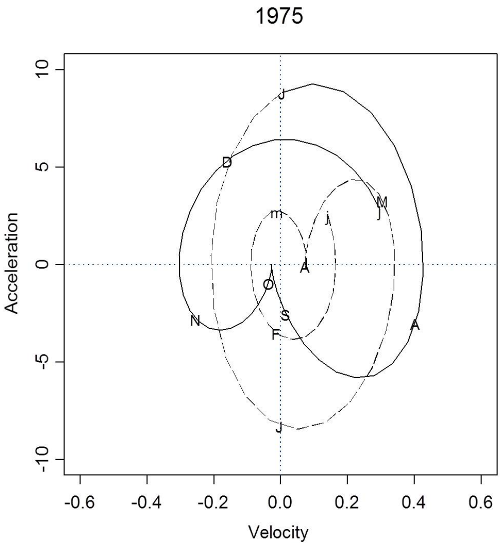
Years of World War II

- In times of war, people don't take holidays, make do with what they have, and spend less at Christmas.
- During World War II, the seasonal cycle became very small, the demand for nondurable goods, like the war itself, was steady throughout the year.

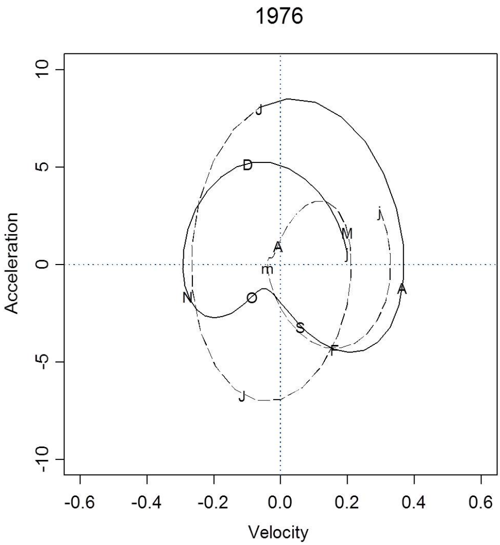
Phase-plane plot of year 1974



Phase-plane plot of year 1975



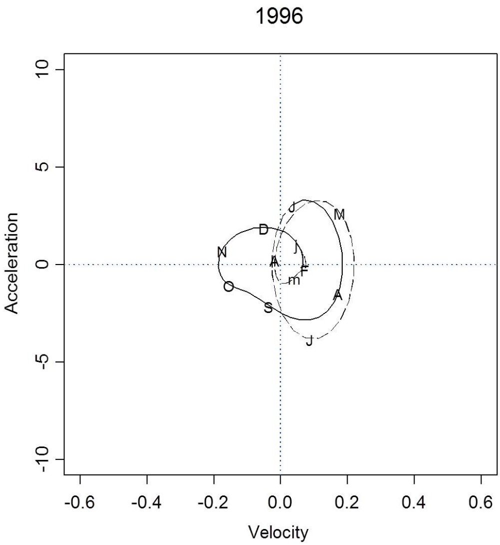
Phase-plane plot of year 1976



Years 1974-1976

- The Vietnam War ended and the OPEC oil crisis happened.
- The fall cycle shrank.

Phase-plane plot of year 1996



Last decade of the 20th century

The size of all the three cycles have become much smaller. Why?

- Is variation now smoothed out by information technology, which does not depend much on manpower?
- Are the aging baby boomers spending less?
- Are personal computers, video games, and other electronic goods really durable?
- Has manufacturing now moved off shore?

Conclusions

Phase-plane plots are great ways to inspect seasonality. Things to look for:

- A substantial cycle.
- Size of the radius: larger radius $\Rightarrow$ more energy transfer in the event.
- Horizontal location of the center: net positive/negative velocity.
- Vertical location of the center: net velocity increase/decrease.
- Changes in the shapes of the cycles from cycle to cycle.